Neural Networks

Statistical Mechanics, and

Tornados

Caren Marzban

National Severe Storms Laboratory, and
Department of Physics, University of Oklahoma

To the memory of Abhijit Kshirsagar (1961-1996)

Some Definitions

Brain: "... a collection of individual entities, called neurons, *interacting* with one another through connections ..."
Kandel, E.R. & Schwartz, J.H. (1991). Principles of Neural Science, 3^{rd} ed. New York, NY: Elsevier Publishing.

"Quantum Brain"
1906 Nobel Prize, Camillo Colgi and Ramon y Cajal.

Society: "An enduring and cooperating social group whose members have developed organized patterns of relationships through *interactions* with one another."
Webster's Ninth Collegiate Dictionary.

Statistical Mechanics: "... to predict the properties of systems composed of large numbers of particles in terms of the properties of the individual particles and of the forces between them."
Encyclopedia of Physics, Addison-Wesley, 1981.

Interacting particles $\longleftrightarrow$ communicating neurons.
Spin-1/2 $\longleftrightarrow$ On/Off neurons (nodes).

Ensemble of interacting things (not necessarily large).

Neural Network (NN).

What's it good for?

Depends on whom you ask.

Neuroscientist?

Physicist?

Statistician?

Me?
Modeling the brain
High energy elementary particles
Predicting TV viewers and nonviewers
Predicting Math scores
Predicting tornados ←—
etc.

But, first, a bit more on NNs...

Memory (Associative)

Spin configuration $\sigma = \{+1, +1, -1, +1, -1, ...\} \sim$ figure:

What weights, ω_{ij}, represent a given spin configuration?
Answer: $\omega_{ij} = \frac{1}{N}\sigma_i\sigma_j$ (Hebb Rule).

Example: Suppose $s_i(t=0) = \begin{cases} -\sigma_i & i = 1, ...n \quad \text{(wrong)} \\ +\sigma_i & i = n+1, ..., N \end{cases}$

$$\begin{aligned} s_i(t=1) &= sgn[B_i(t=0)], \\ &= sgn[\sum_j^N \omega_{ij} s_j(t=0)] \\ &= sgn[\sum_j^N \frac{1}{N}\sigma_i\sigma_j s_j(t=0)] \\ &= sgn[\frac{1}{N}\sigma_i \left(-\sum_{j=1}^{n} \sigma_j^2 + \sum_{j=n+1}^{N} \sigma_j^2 \right)] \\ &= sgn(1 - \frac{2n}{N})\sigma_i \\ &= \sigma_i \quad \text{if} \quad n < \frac{N}{2} \end{aligned}$$

Reorganize Smiley, if not too noisy.

Brain in a Heat Bath

$\omega = + \rightarrow$ Ferro
$\omega = \pm$ & regular $\rightarrow$ Antiferro
$\omega = \pm$ & random $\rightarrow$ Spin Glass

Stochastic: $P[s_i(t+1) = +1] = f(B_i(t)) = \frac{1}{1+\exp^{-\beta B}}$.

$$H = -\frac{1}{2}\sum_{i,j}^{N} \omega_{ij}\ s_i\ s_j \quad , \quad Z = Tr_s \exp^{-\beta H}\ .$$

Example:

$$\begin{aligned} < s_i > &= (+1)f(B_i) + (-1)f(-B_i) = tanh(\beta B_i) \\ \text{If} \quad \omega_{ij} &= \frac{1}{N} \quad (s'_i = \sigma_i\ s_i \rightarrow \omega'_{ij} = \sigma_i \omega_{ij} \sigma_j = 1/N) \\ < s_i > &= tanh(\beta \frac{1}{N} \sum_{j}^{N} < s_j >) \\ < s > &= tanh(\beta < s >) \end{aligned}$$

Critical T (Curie Point)
Phase Transition.
Above T $\longrightarrow$ Amnesia.

Layered NNs

Motor-Visual portions of the brain are layered.

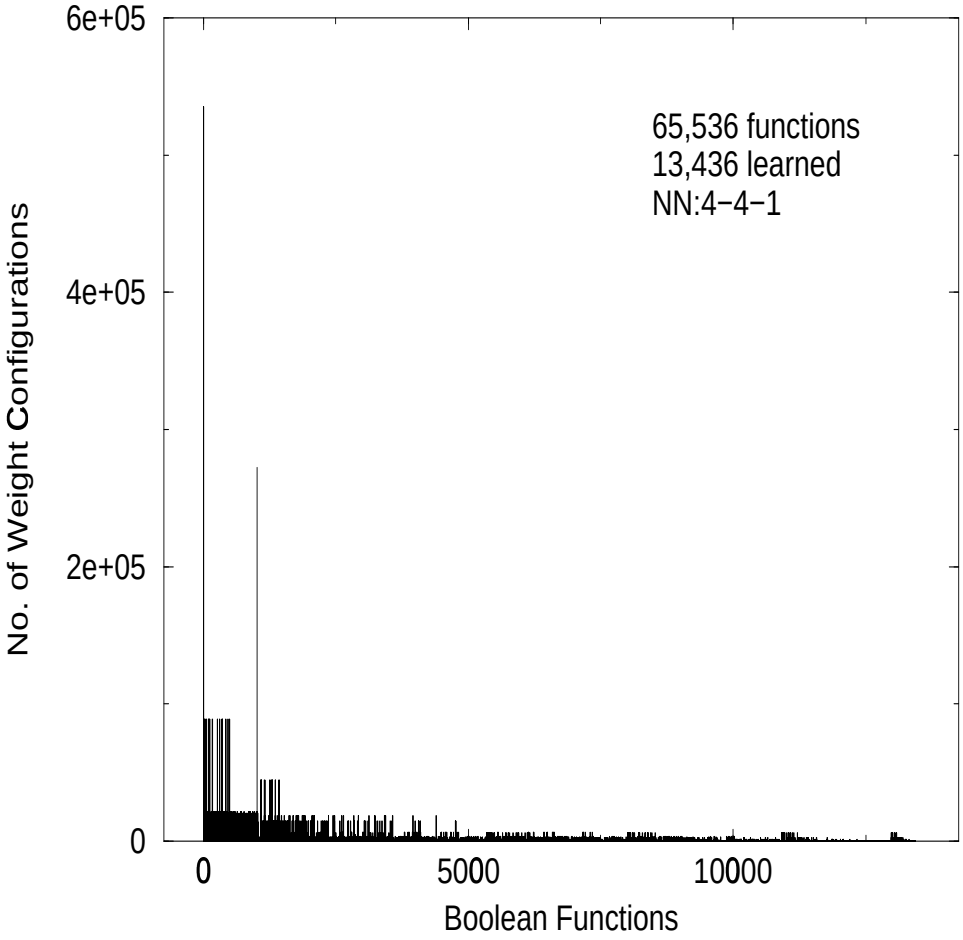
Good for learning complex **functions**, $y = F(x_1, x_2, ...)$.

Example: Name-face mapping.

Example: Boolean (binary) functions.
$2^{2^{N_{in}}}$ simple, complex functions. $N_{in} = 2$:

x_1	x_2	y								
0	0	0	0	0	0	1	...	0	...	1
0	1	0	0	0	1	0	...	1	...	1
1	0	0	0	1	0	0	...	1	...	1
1	1	0	1	0	0	0	...	0	...	1

Which functions can be "learned"?



No. of weight (± 1) configurations representing a given function $\sim$ complexity (probability, P_f) of that function.

Given an architecture, some functions are easy, some are hard, and some are unlearnable.

• Wired-in.

Diversity of the allowed functions $\sim$ Entropy

$$S = \sum_f P_f \log P_f.$$

Fact: Among all the alternatives, those with larger entropy are probabilistically preferred.

• Learning is a direct consequence of the 2^{nd} Law!

Enough (amazing) formalities.

Anything practical?

Tornado Prediction

$\exists$ a function mapping tornado attributes to existence (0/1).

Attributes: Base, depth, shear, gate-to-gate velocity difference, ...

What is that function?

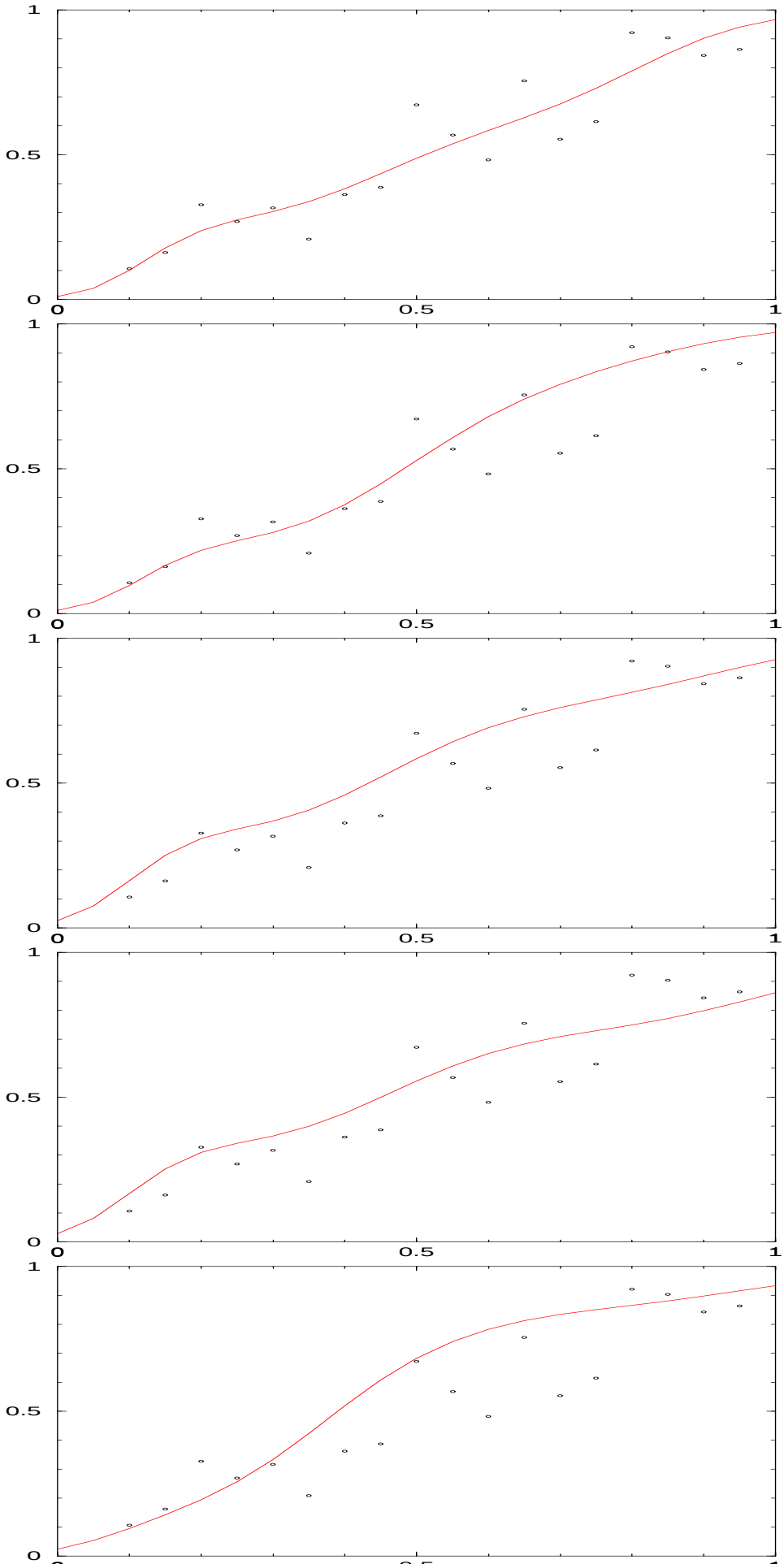
Collect lots of data: inputs (attributes) and outputs (0/1).
Start training (learning), i.e. finding the "correct" ω's.

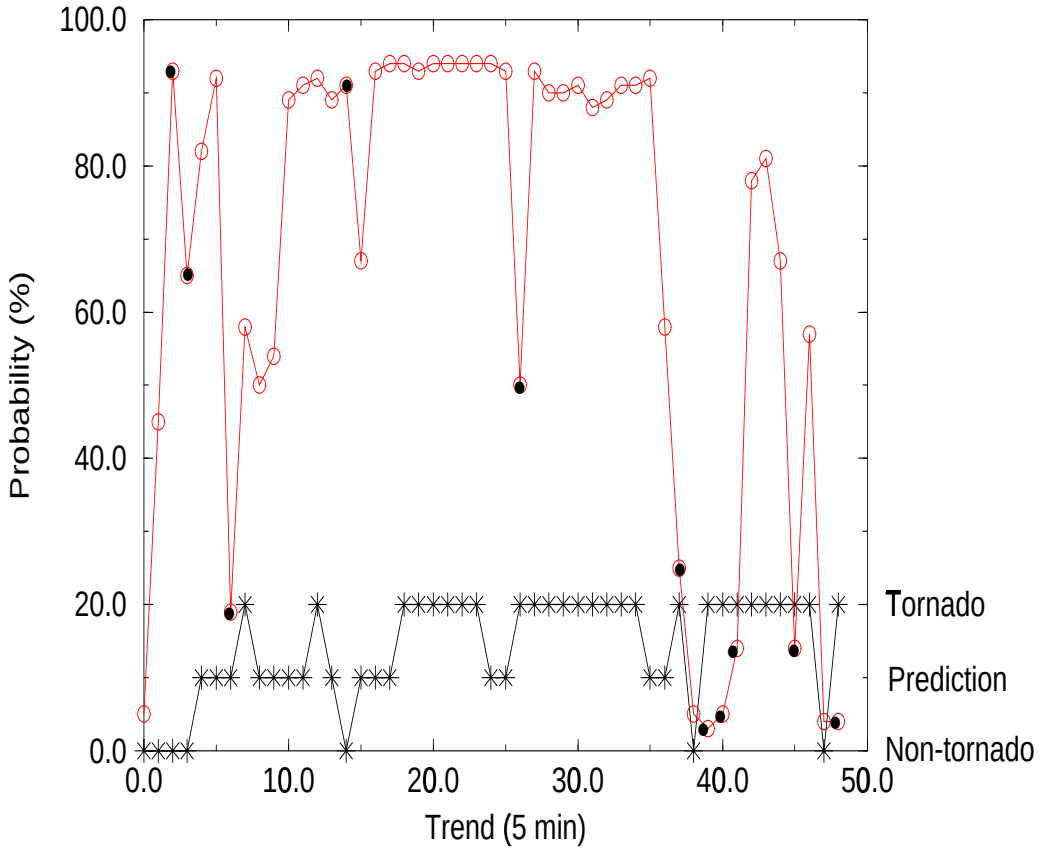
. . . . . Nine Months Later . . . . . .

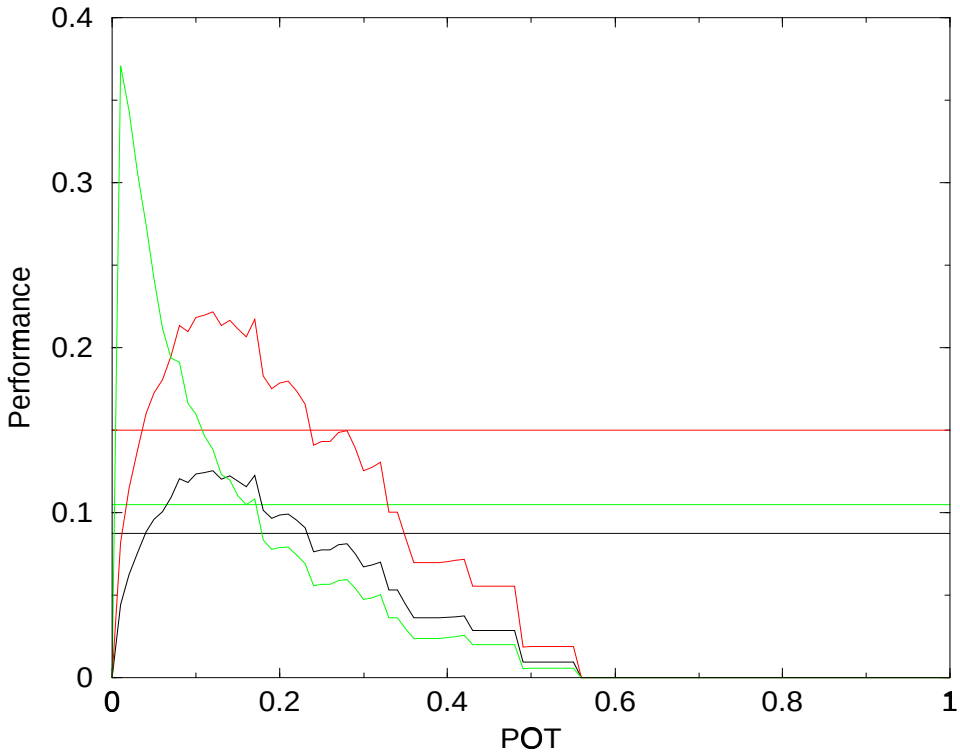
. . . . . NN is born . . . . . .

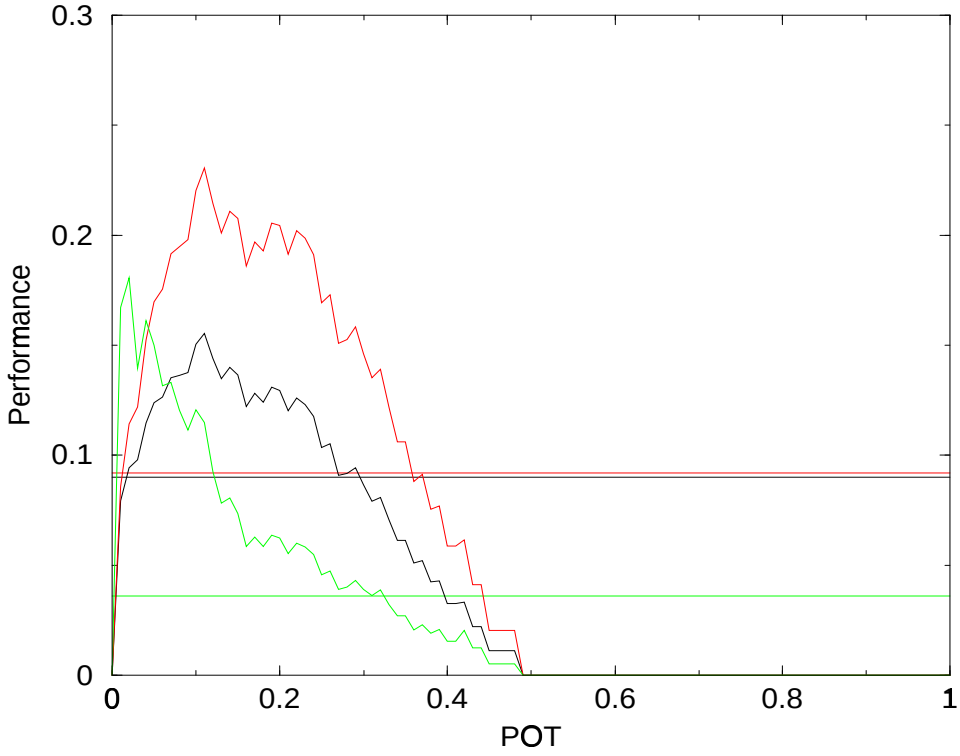
How does it perform?

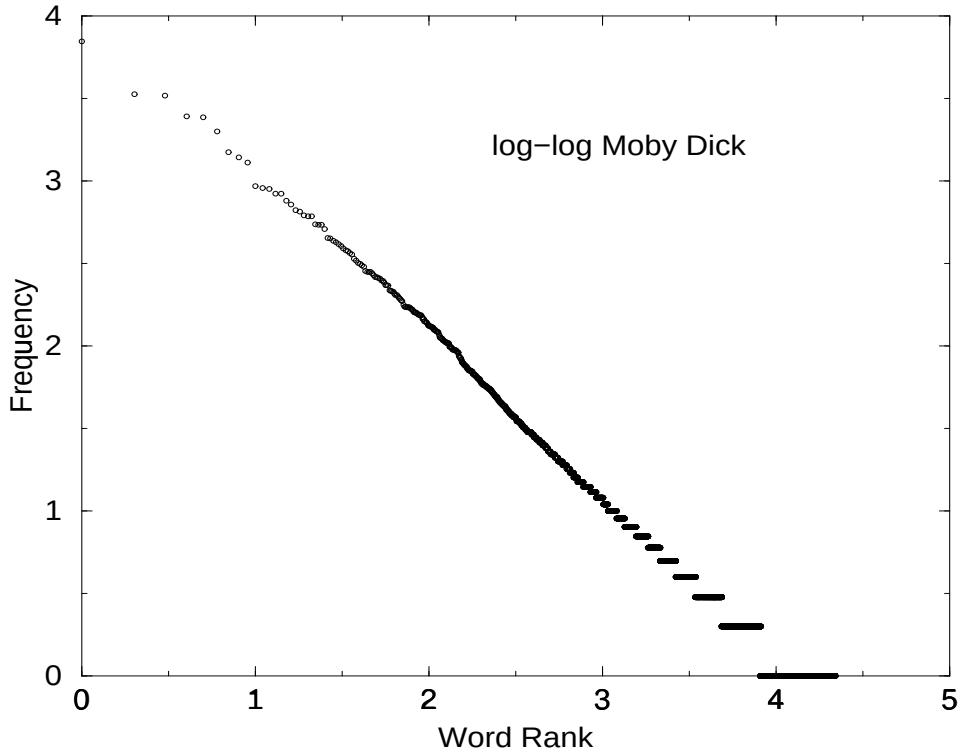
Figure 1: Robustness. (121 weights) No cuts, 20 cuts, 40 cuts, 60 cuts, 100 cuts.




100.0
80.0
60.0
40.0
20.0
0.0
Probability (%)
0.0
10.0
20.0
30.0
40.0
50.0
Trend (5 min)
Tornado
Prediction
Non-tornado


0.4
0.3
0.2
0.1
0
Performance
0
0.2
0.4
0.6
0.8
1
POT


0.3
0.2
0.1
0
Performance
0
0.2
0.4
0.6
0.8
1
POT



Conclusion: Deep Thoughts

What happened to the brain?

Language is a manifestation of the brain.

Moby Dick and Zipf's (power) Law.

Self-Organized Criticality (SOC) and the sand-pile.

Meaning: Universality, scale-free, ∞ correlation lengths, etc.

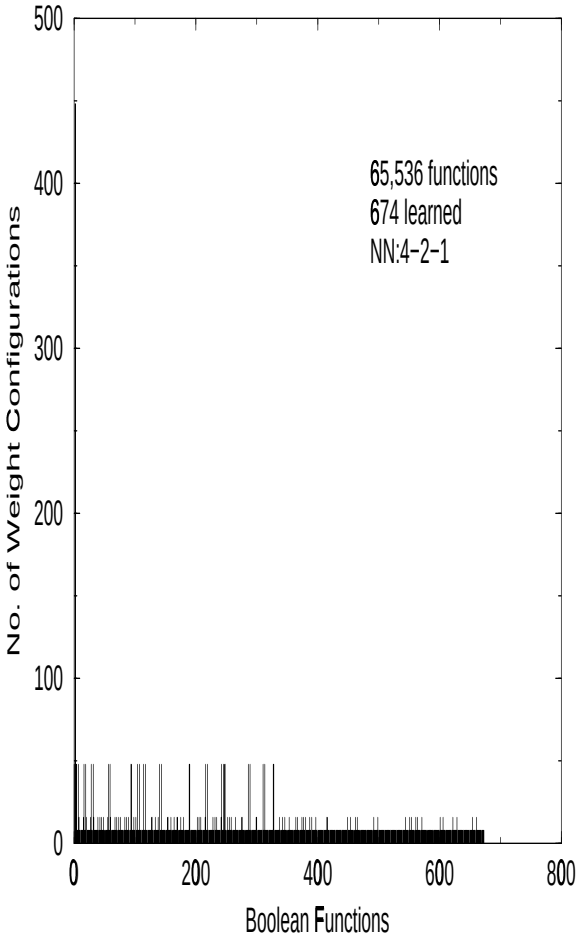
Wait! It gets deeper.

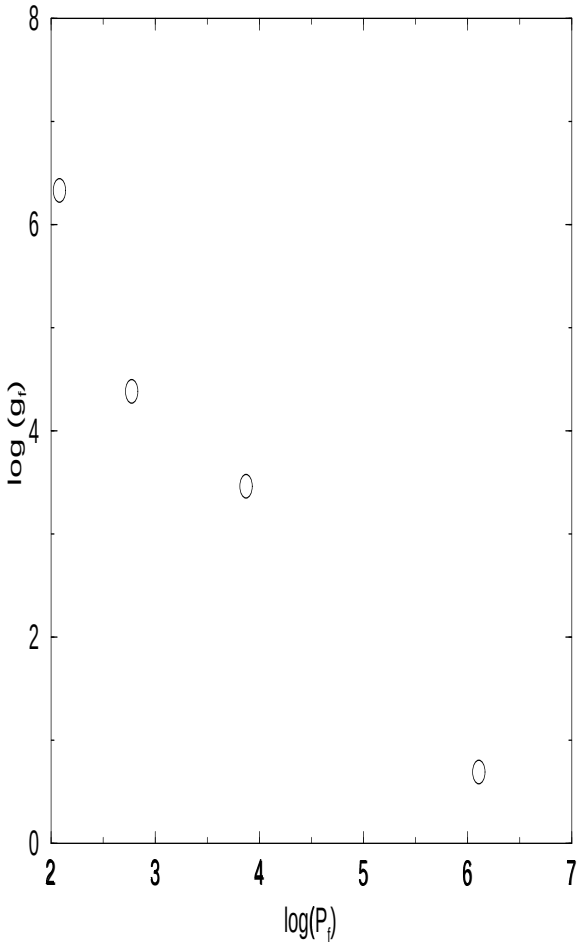
Brain
Earthquakes
Biological Systems
Species Extinction
Social Systems
Economic Systems
Wars

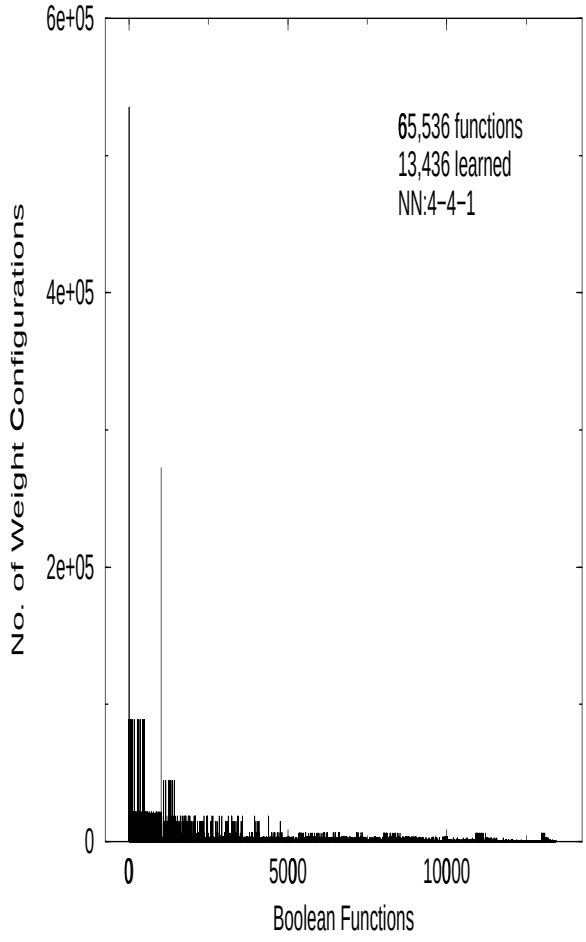
All have NN structure.
And SOC is inherent in NN complexity!!

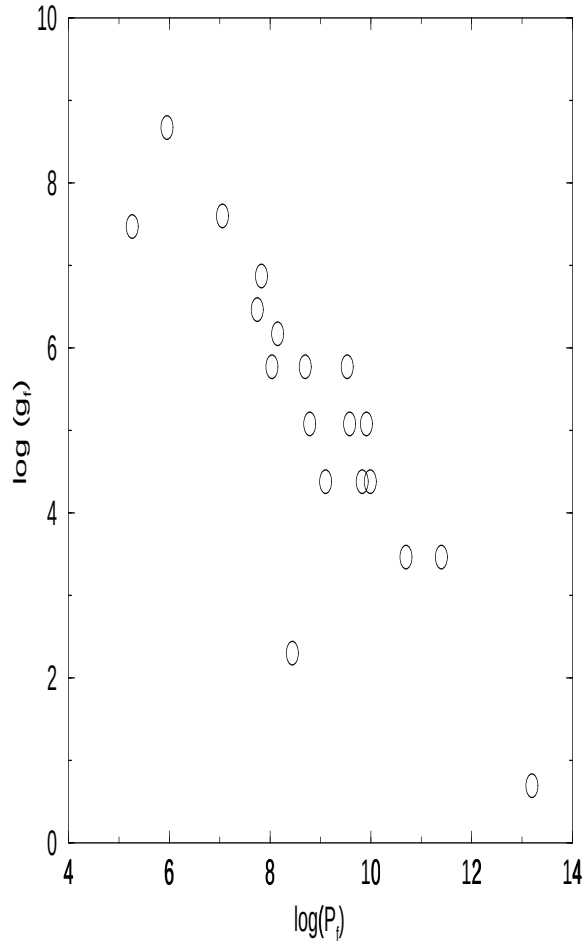
$$1 = \sum_f P_f = \sum_{f'} g(f') P_{f'}$$

It's all wired-in!


65,536 functions
674 learned
NN:4−2−1
No. of Weight Configurations
Boolean Functions
0
100
200
300
400
500
0
200
400
600
800


log (g_f)
log(P_f)
0
2
4
6
8
2
3
4
5
6
7


65,536 functions
13,436 learned
NN:4−4−1
No. of Weight Configurations
Boolean Functions
0
2e+05
4e+05
6e+05
0
5000
10000


log (g_f)
log(P_f)
0
2
4
6
8
10
4
6
8
10
12
14